DOD-Explainer: Explainable Drug Overdose Deaths Predictor from Crime and Socioeconomic Data

Abdulaziz Alhamadani*, Shailik Sarkar*, Lulwah Alkulaib* ‡, and Chang-Tien Lu*
* Department of Computer Science, Virginia Tech, Falls Church, VA 22043 USA
‡ Department of Computer Science, Kuwait University, Kuwait
{hamdani, shailik, lalkulaib, ctlu}@vt.edu

2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10021054

***Abstract*—Drug overdose deaths are a dreadful crisis that leads to substantial societal impairments. Its harmful impact directly affects families and communities. To assist policymakers in mitigating this crisis, it is crucial to study the societal, economic, and criminal contributing factors linked to the crisis. Unfortunately, current data-driven works assume a singular factor, such as poverty being the cause and disregarding other realistic causes. Besides, recent works exhibited a lack of explainable models and spatial analysis of the crisis. Thus, DOD-Explainer links the gap by developing a realistic framework that predicts highly impacted counties of drug overdose deaths from crime and socioeconomic data. DOD-Explainer overcomes the challenge of data scarcity by proposing three data augmentation methods. Then, an algorithm is proposed to provide realistic explanations of the leading causes of the crisis. The results show that our application achieves the best predictive accuracy from several models, accurately identifies the most/least impacted counties by the crisis, and reveals the most contributing factors of drug overdoses.**

***Index Terms*—Drug overdose deaths, crime, socioeconomic, prediction, spatial analysis, data augmentation, explainability**

I. Introduction

The number of drug overdose deaths in the U.S. was reported to be more than 100,000 people between 2020 and 2021. According to the Centers for Disease Control and Prevention (CDC), the overdose crisis is very serious. Every day, 136 people lose their lives due to only an opioid overdose. Drug overdose deaths or drug-induced deaths include all deaths caused by either intentional or unintentional acute poisoning by drugs (illicit or street drugs) and medical conditions resulting from chronic drug use. Further, the deaths involving illegally manufactured synthetic opioids, for example, fentanyl and other stimulants, are increasing by an alarming number in the past few years. For instance, the total number of drug overdose deaths from 2017 to 2020 in Arizona and California nearly doubled in four years. As researchers, our mission lies in providing a better understanding and response to this crisis by collecting, analyzing, and identifying the regions that need assistance so they can acquire the needed support from authorities.

There are many statistical attempts to analyze the drug overdose crisis in the U.S. Some studies focused their analysis on race [1], [2], but race is not the actual factor that contributes to the drug overdose crisis, and there are factors that require attention, such as socioeconomic factors. For instance, these studies delved into more specific analyses and showed that drug overdosing was concentrated in areas of high rates of unemployment [3], poverty, low income, and low education [4]. In addition, the demographics of a population in an area, such as the age group and families in economic distress [5] have shown to be factors for the crisis. More evidence explains the relationship between the drug overdose crisis and the combined socioeconomic factors [6], [7]. Therefore, it is crucial to include those factors in analyzing the drug overdose crisis.

Socioeconomic factors are not the only explanation for the crisis. A report from the U.S. Department of Justice reveals the association between the drug-using lifestyle and the crime problem. It shows that a person who uses drugs is more likely to commit crimes than a nonuser in the general population. Moreover, a range of 38% to 85% of arrestees tested positive for any drug use. Additionally, this drug lifestyle often pushes offenders to commit crimes to support their drug habit [8]. Contrerase et al.'s [9] study concluded that drug activity increases assaultive violence and crime rates. Certainly, there is a close relationship between drug prevalence and crime [10]. Subsequently, our study combines the socioeconomic factors and crime rate data to provide explanations for the crisis and identify high-risk areas for drug overdose deaths.

Current studies have positively contributed toward understanding drug abuse and its prevalence, yet the field exhibited a lack of work that explains the crisis, provides spatial analysis and forecasts the high-risk areas of a drug overdose. Most works focused on analyzing the trends of drug abuse on social media or through limited samples of questionnaires [11]–[23]. Few works examined drug abuse behavior and provided explanations through a small subset of groups and reasons, for example, adolescents [24], [25] or homeless groups [26]. Consequently, considering the existing works in the field and the drug overdose crisis, the exigent reasons require an explainable and spatial analysis of the matter.

To address those challenges, we propose **D**rug **O**verdose **D**eaths **Explainer** (DOD-Explainer), a novel (framework that provides a county-level spatial prediction of drug overdose deaths from crime and socioeconomic data, solves the scarcity of data, and offers interpretations to mitigate the crisis). In this work, we seek to link the gap in the data-driven drug overdose crisis literature through the following tasks: **(T1) Find effective solutions to the scarce and heterogeneous data.** Our

framework combines six categories of socioeconomic factors of 7 varying U.S states, their crime data, and drug overdose mortalities based on the county-level. To create enough data to train the machine learning and deep learning models, we propose solutions to augment the heterogeneous data in three different methods. **(T2) Develop spatial prediction application of high-risk areas of drug overdose deaths.** The framework applies various Machine Learning (ML) and Deep Learning (DL) methods to identify the impacted counties by the crisis and demonstrates several case studies. **(T3) Provide an explainable framework for the crisis.** There are many factors that play an important role in the drug overdose crisis, such as crime rates, age demographics, education levels, economics (family and individual income), occupation types, poverty levels, and families receiving Supplemental and Nutrition Assistance Program (SNAP) benefits. This work identifies the major factors for this crisis, so decisions can be made by officials in response to the given explanations. The main contributions of our work are summarized as follows:

- **Develop a realistic framework that spatially forecasts highly impacted U.S. counties of drug overdose deaths.** The framework considers realistic factors of the drug crisis by utilizing real-world contributing factors such as crime data and socioeconomic statistics to predict the areas of drug overdose mortalities.
- **Propose effective solutions to scarce and heterogeneous data**. The field of drugs is typically discrete and sensitive. Therefore, the data is scarce to train ML and DL models. Therefore, we augment the data with three different methods. Our framework also combines data from multiple resources based on each county in the U.S to simulate realistic factors of the crisis.
- **Provide an explainable algorithm of realistic interpretations of the crisis.** There are many socioeconomic and crime factors that contribute to the crisis. DOD-Explainer interprets its predictions to domain experts to determine key factors through an algorithm we developed to rank the most contributing factors to the crisis.
- **Conduct extensive experiments to demonstrate the effectiveness of the proposed framework.** The proposed framework was performed on two tasks (classification and regression). Then, six experiments and 3 case studies are conducted. The in-depth experiments demonstrate the superiority of the work on a wide variety of data augmentation methods and a comprehensive selection of realistic factors for the crisis.

II. Related Works

This section introduces related work in several research areas.

Drug overdose crisis modeling and forecasting: Despite the positive contributions that recent studies have made to understanding drug abuse and its prevalence, the field of drug abuse research has shown a shortage of work that explains the issue, provides spatial analysis, and predicts the high-risk locations of a drug overdose. Most studies analyze drug misuse trends on social media or use small questionnaire samples. Some existing works explored drug addiction behavior and offered explanations and identifying factors, yet some of those works only highlighted one type of drug user [24], [26]. The majority only analyzed drug use and abuse patterns [11], [12], [14], [15], [17]. Further, others have studied social media platforms, including Instagram, Twitter, and YouTube, to monitor and identify those patterns [19]–[21], [27]. Even though those initiatives significantly and positively impacted important parts of current drug use and misuse problems, yet they also presented some significant challenges mentioned in section I.

The relatively static features of the geographic regions that are crucial to the propagation of addiction are the subject of very few existing works. As a result, this study forecasts drug overdoses in a novel way using socioeconomic factors and crime reports data through county-level spatial analysis. Socioeconomic status, crime, and addiction have always been related [28], [29]. People belonging to higher groups in the social hierarchy often have access to better care than those at the bottom. Thus, socioeconomic level differences are known to affect drug usage rates. Moreover, Mukherjee et al. [30]conducted a case study on Connecticut to predict short-term drug overdose deaths using internet search trends. Since addiction is under-reported and can be vague to discern, existing addiction numbers aren't always accurate [12], but this study proposes data augmentation solutions to enhance that. Further, Gong et al. [17] presented a model that effectively predicts the short-term trend of opioid abuse, but their work does not identify high-risk areas associated with drug overdoses, and their model does not provide an explainability component as we propose in our work.

Explainability for drug overdose crisis: Explainability techniques are essential to understanding how models operate and provide insight into whether the results presented by the model were correct. Explainability techniques are applicable to various applications and models like classification, prediction, and image segmentation [31]. Ying et al. [32] proposed GCN explainer, the first general model-agnostic approach to provide interpretable explanations for predictions of any GCN-based model. DeepCOVIDExplainer [33] presented an explainable deep neural network method for the automatic detection of COVID-19 symptoms from chest radiography images. Rodriguez et al. [34] provided exploratory analysis modules, enabling interpretation of their COVID-19 forecasts. Providing model explainability brings more confidence in results from ML and DL models. Inspired by these works, we include an explainability module in our work to bring transparency to our predictive analysis.

III. Methods

This section describes the methodology in three parts. The main objective is to forecast the drug overdose deaths in a county given a combination set of crime reports and socioeconomic features dataset illustrated in Table I. First, we use traditional ML models to predict drug overdose deaths on

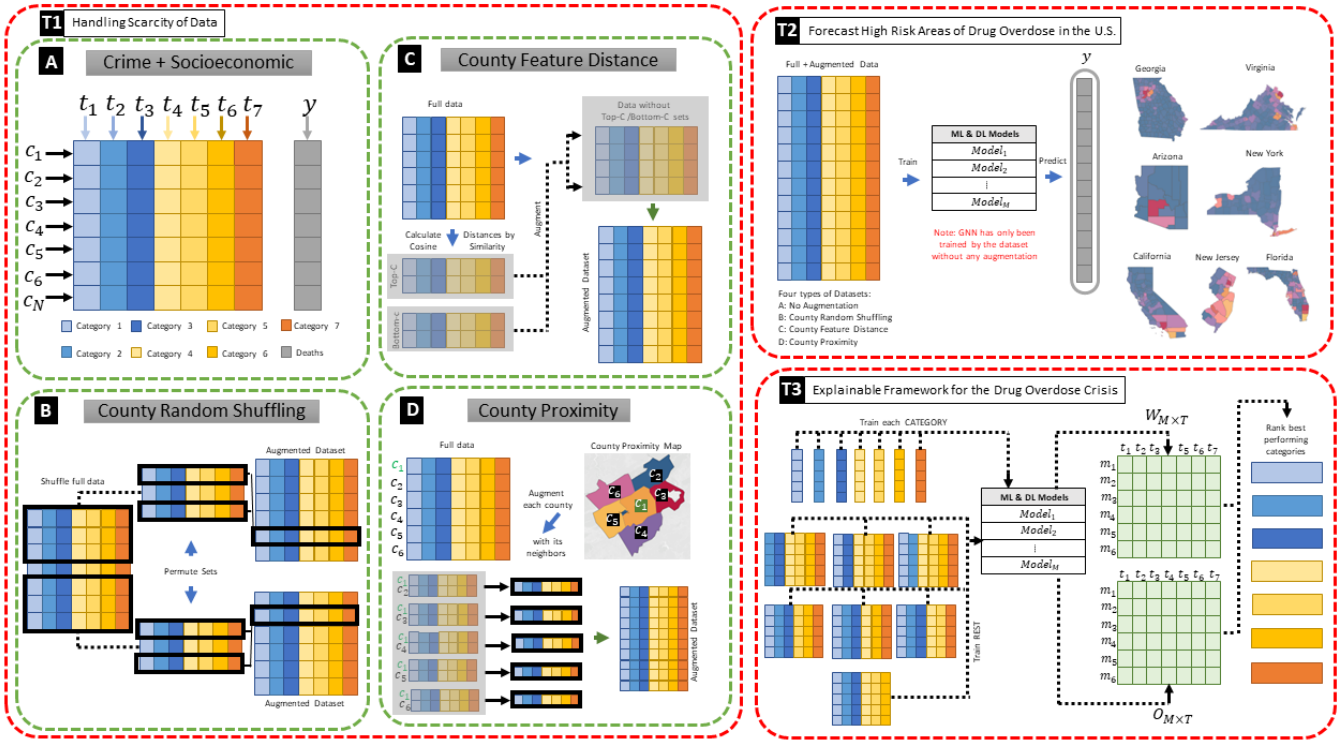

Fig. 1: All tasks are illustrated in this figure. $T1$ shows the four types of datasets: (A) Dataset without augmentation, (B) County random shuffling, (C) County Feature distance, and (D) County Proximity augmentation. $T2$ displays the forecasting approach based on $T1$. $T3$ describes the explainability according to section 3

a county-level basis. Second, to solve the issue of data paucity at the county-level, we enhance the data by proposing three data augmentation methods. Then, we implement DL models, including a Graph Convolutional Network (GCN) model without data augmentation to leverage the spatial correlation among different counties. Finally, we introduce the explainability method to highlight the categories that contribute the most/least to the crisis.

A. Data Augmentation

The dataset collected regarding drug overdose deaths is very sensitive and highly anonymous to protect the identity of the users. Obtaining fine-grained information, for example, based on the zip-code level is not available for many U.S. states because most of the publicly reported numbers are suppressed to protect the privacy of individuals and families. Therefore, we seek to utilize county-level data to provide predictions of drug overdose mortalities for risk evaluation of spatial areas. Fortunately, collecting county-level datasets is obtainable but comes with a cost. The number of counties in some states is very small, which in return will negatively affect the prediction results for traditional ML and some DL models. Therefore, we propose three different data augmentation methods to solve the paucity of data.

T1. Data augmentation:(fig 1) When insufficient data is available or suppressed, such as drug overdose deaths, data augmentation makes collecting new data dispensable. Data augmentation is an approach that expands the diversity and size of training data from the original data. In addition, it has proven its ability to enhance models' forecasting accuracies and to improve the models' robustness toward overfitting. [35]–[37].

TABLE I: Description of the dataset categories, the number of features/factors for each category, and an example.

Category	No.	Example
Age Demographics	21	20 to 24 years
Crime Reporting	6	Murder
Drug Overdose	1	Deaths count
Education Attainment	19	Less than high school
Food Stamps SNAP	25	Families getting food stamps
Household Income	11	$15,000 to $24,999
Occupation Types	8	Service occupations
Poverty Status	8	200 percent of poverty level

1) County-Random-Shuffling Data Augmentation: The first applied method is *permutation* [35], [37] and random *shuffling* [38] data augmentation, which mimics random-county data. This method showed improvements in inaccuracies of other models and can be further extended for any level of a geographical area such as zip code [39]. For example, a new synthetic data point where the mapping between the features and label (drug overdose death) is generated when the data of

two counties are given.

The permutation method for data augmentation was proposed by Um et al. [37], and Eyobu et al, and [38] applied the shuffling approach. In shuffling, the data(county) points are randomly rearranged. While in permutation, the dataset is divided into equal segments, then each segment is permuted. This work applied both methods (random shuffling and permutation). The following details explain how the full dataset is augmented. CS_N^T indicates the full dataset in the shape of a matrix where T is the number of columns. The matrix columns $CS_i^T = \{x_i^1, x_i^2, ...x_i^T, y_i\}_{i=1}$ are the heterogeneous data of the combination of crime rate data and socioeconomic features, and the last column is the drug overdose deaths. The rows denoted as N are all the counties.

$$CS_N^T = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^t & y_1 \\ x_2^1 & x_2^2 & \dots & x_2^t & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^t & y_n \end{bmatrix}$$

At first, random shuffling is applied, then the rows of the matrix CS_N^T are divided into k equal-sized segments for the permutation. Each segment's size is determined by $\frac{n}{k}$, and this results into the segmented matrix $CS_{\clubsuit}^T$

$$CS_{\clubsuit}^T = \begin{bmatrix} CS^T(1) \\ CS^T(2) \\ \vdots \\ CS^T(K) \end{bmatrix}$$

where

$$CS_{\clubsuit}^T(i) = \begin{bmatrix} x_{(i-1)\frac{n}{k}+1}^1 & x_{(i-1)\frac{n}{k}+1}^2 & \dots & x_{(i-1)\frac{n}{k}+1}^t \\ x_{(i-1)\frac{n}{k}+2}^1 & x_{(i-1)\frac{n}{k}+2}^2 & \vdots & x_{(i-1)\frac{n}{k}+2}^t \\ \vdots & \vdots & \dots & \vdots \\ x_{i\frac{n}{k}}^1 & x_{i\frac{n}{k}}^2 & \dots & x_{i\frac{n}{k}}^t \end{bmatrix}$$

The rows are permuted for each $CS_{\clubsuit}^T(i)$ ($i \in [1, k], k = 10$) and each permutation engages two rows (counties') features to augment the data. During permutation, the features of each county are augmented using the augmentation function.

Augmentation function: Given the row of a county $x_{(i-1)\frac{n}{k}+1}^1$ and $x_{(i-1)\frac{n}{k}+2}^1$, the augmentation is defined as $x_{i-j} = F(x_i, x_j)$ where F is the augmentation function (e.g. average, maximum). Let x be the feature such as burglary. The augmentation function can be represented as:

$$F(x_i, x_j)[burglary] = AVERAGE(x_i, x_j) \quad (1)$$

Along with the county-random-shuffling data augmentation method, we propose another two data augmentation methods to conduct the study. County-random-shuffling method examines if random and unrelated data augmentation can enhance the prediction of drug overdose deaths ratio.

2) County-Proximity Data Augmentation: This data augmentation method relies on the information of the target area's neighboring (surrounding) regions. For example, the surrounding areas of Fauquier County in Virginia are Clarke, Culpeper, Loudoun, Prince William, Rappahannock, Stafford, and Warren counties. Spontaneously, it is questionable to generate data from similar or neighboring regions because that will create bias. However, we are investigating the assumption called "the neighborhood effect" [40]. The concept indicates that the social science and economic factors that polarize neighborhoods may directly or indirectly affect an individual's behavior. Further, neighboring regions can have direct or indirect effects. In this work, no assumptions are made that there is a direct effect between a county's socioeconomic and crime data on a neighboring one. Instead, we are investigating this concept in the experiment section through County-Proximity Data Augmentation.

Algorithm 1 County-Proximity Data Augmentation

Input: Crime-socioeconomic CS_N, County-Adjacency-Map CAM_M
Output: Augmented County-Proximity Set CD
Initialize: County-Adjacency Matrix $AM_{N\times N}$

```
Build AM_{N×N} from CAM_M
for each row am_i ∈ AM_{N×N} do
    for each column am_j ∈ AM_{N×N} do
        if (am_i, am_j) = True then   ▷ They are neighbors
            tempCS_1 = Find am_i from CS_N
            tempCS_2 = Find am_j from CS_N
            Augment(tempCS_1, tempCS_2) equation 1
            update CD
        end if
    end for
end for
Return CD
```

In the exhibited pseudo-code 1, the input comprises of two parts. The first is the crime data combined with the socioeconomic factors CS for all the available counties N. The second input is the list of county adjacency data CAM_M obtained from the U.S. Census Bureau[1] where M is the number of all U.S. counties. The output is the set of County-Proximity augmented data notated as CD. At first, we filter the counties of only the states we collected data from to match the number of counties from M to N. Then, the counties adjacency matrix $AM_{N\times N}$ is built based on CAM_M, and of size $N \times N$ where N is the number of all available counties in CS_N. The algorithm searches $AM_{N\times N}$ to find if there is a relationship between one county and another. If they are neighbors, then both counties will have their data augmented to generate a new synthetic county. Each time a new one is generated, CD is updated. Once the search is complete, CD is employed in the experiment for all the models except for the GCN model because the graph already connects the spatially related counties.

[1]https://www.census.gov/geographies/reference-files/2010/geo/county-adjacency.html

3) Feature-distance Data Augmentation: This method is different from the previous two because it aims to create synthetic county data from the tails of the distribution, either the highest or the lowest. The previous methods could be either generating random data, closely related data, or disparate data. Therefore, we propose feature-distance data augmentation. This method calculates the distance among all counties and finds the top-C counties with the greatest distance. Then, it generates synthetic data. This method does not only consider the top-C because relying only on the top will create an imbalance in the augmented dataset, which may subsequently impact the performance of the prediction. Thus, the method also calculates the bottom-C counties with the least distance to create balanced data and includes the distribution's middle and tails.

Algorithm 2 Feature-distance Data Augmentation

Input: Crime-Socioeconomic CS_N
Output: Augmented Feature-Distance Set FD
Initialize: h = 0, j = 0, Temp-copy Crime-Socioeconomic cCS_Z = CS_N ,Top-counties T_C, Bottom-counties B_C

```
while h < N do
    h = h + 1
    while j < Z do
        j = j + 1
        if h ≠ j then
                                        ▷ calculate cosine similarity
            temp = CosineSimilarity(CS_i, cCS_j)
            if temp > any element in T_C then
                update T_C
            else if temp < any element in B_C then
                update B_C
            end if
        end if
    end while
end while
For each element in CS_N not in T_C or B_C:
FD ←Augment each element from CS_N with each element in T_C and B_C  1
Return FD
```

The details of this method are exhibited in pseudo-code 2. The input of this algorithm is only the same unaugmented data which is the crime and socioeconomic data CN_N. The output is the augmented set of feature-distance FD. Then, two sets of data are initialized to store the counties with the greatest distance T_C and the bottom counties with the least distance B_C. C stands for the number of top or bottom counties. The variable C relies on the number of counties in each state; for example, the state of Georgia has more than 150 counties, and C in this case will be 15. Therefore, C varies from 3, 5, 10 to 15 depending on the number of counties within a state. At first, the algorithm calculates the distance between each data point $\in CS_N$ and all the other points $\in cCS_Z$ (which is a copy of the same table of the same size). For example, the data of a county in Virginia $CS_i^v = \{a_i^v, b_i^v, ..., z_i^v\}_{i=1}^N$ is measured by the cosine similarity against all the other counties in cCS_Z such as $cCS_j^v = \{a_j^v, b_j^v, ..., z_j^v\}_{j=2}^Z$ (see equation 2). The sets T_C and B_C are updated whenever the $temp$ value satisfies either condition. Once all the distances are calculated, the augmentation process starts.

$$\cos(\mathbf{CS_i}, \mathbf{cCS_j}) = \frac{\mathbf{CS_i} \cdot \mathbf{cCS_j}}{\|\mathbf{CS_i}\|\|\mathbf{cCS_j}\|} \tag{2}$$

For each element in CS_N and also not available in either T_C nor B_C, the algorithm augments each element with all elements in T_C and B_C based on equation 1. Each time two data points are augmented, FD is updated, and FD will return the full feature-distance augmented set for the experiment. The three methods of data augmentation in this paper offer more training data to the models and enhance the models' abilities to predict drug overdose deaths. Following are the models used on both unaugmented and augmented data to predict drug overdose deaths.

B. Machine Learning and Deep Learning Models

T2. Spatial prediction of high-risk areas of drug overdose deaths: The problem of predicting drug overdose deaths is considered in two approaches. The first approach teaches the models how to predict drug overdose death ratios per 100,000, then ranks the high-risk counties within each state. The second approach considers the problem as a binary classification where the threshold is the national opioid overdose mortality rate of 12.5 in 2017 [41]. The models used for both approaches are as follows:

a) Ridge Regression: Ridge is a linear model with L2 regularization, ideally enhancing the basic linear regression model. L2 regularization improves in providing a larger coefficient to the most significant features.

b) Lasso Regression: It is a regularization method used for prediction as Ridge regression, but Lasso differs by shrinking the coefficients to zero for normalization.

c) SVR: This is a form of SVM proposed for regression evaluation. In our work, we used SVR for prediction and SVM for classification. For the binary classification, the SVM classifier learns to build a decision boundary for lower and linear data and high dimensional and non-linear ones. For prediction, one needs to consider SVR for a solution to the regression problem.

d) LightGBM: It is a sub-form of Gradient Boosting Decision Tree, and works with efficiency and robustness for a large dataset. In our case, the augmented datasets ranged from 17000-24000 data points, making lightGBM an ideal candidate for the prediction problem.

e) XGBoost: Extreme Gradient Boostingis a scalable application of Gradient Boosting Decision Trees (GBDT), which enhances the performance and accuracy of prediction of decision trees such as Random Forest.

f) KNN: This paper utilizes k-Nearest Neighbors algorithm for the second problem, which is the binary classification. SVM is considered a parametric method and can deal with linear or nonlinear boundaries based on the Kernel

function. On the contrary, KNN is a nonparameteric algorithm and can adjust to nonlinear boundaries as the training data increases.

g) MLP: Multi-Layer Perceptron is one of the DL methods we applied for the two problems: binary classification and prediction. For the former problem, MLP can create a better boundary for classification, and for the latter, we adapted the algorithm to perform regression depending on the activation function.

h) Graph Convolutional Network Model: The spatial dependency has a strong effect on the degree of risk [42] when employed, and could improve the estimations accuracy of the risk associated with the counties. Graph Convolutional Network (GCN) is an emerging method that combines node attributes and graph structure. GCN has demonstrated its efficiency in handling spatial dependencies [43].

The problem is formulated as an transductive semi-supervised regression problem on an undirected graph of the unaugmented data. We annotate the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with n nodes. The nodes $\mathcal{V}$ represent the counties, and the edges $\mathcal{E}$ represent the spatial dependency relationships. The edges can be represented in the adjacency matrix: $\mathcal{A} = [A_{ij}] \in \{0,1\}^{n \times M}$. We hypothesize that the closer a county is to another, the higher the spatial dependency will be. Thus, edges are assigned based on the distance between two counties. The distance is calculated based on haversine. we construct the graph by connecting each node to its closest "n" nodes. We attempt to learn the value of n by initializing it at 10 and then decreasing it to achieve sparsity. Finally, we fix the value of n at 5 $X \in \mathcal{R}^{n \times M}$ represents the T features described in Table I.

Once the graph is constructed, the node-level information is enabled to propagate based on the neighborhood. The propagation rule of one layer is defined as the following:

$$g_\theta * X \approx (\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}) X \Theta \tag{3}$$

This is parameterized with Θ, with $\tilde{A} = A + I$, where I is the identity matrix and $\tilde{D}$ is the diagonal node degree matrix of $\tilde{A}$. By stacking two layers of the graph convolutions, our final model is defined as follows:

$$Y = F(g_{\theta_2} * \sigma(g_{\theta_1} * X)) \tag{4}$$

Here, F is the linear layer for the regression task, and σ is the activation function.

C. Drug Overdose mortalities Explainability

T3. Provide an explainable framework for the crisis. Offering predictions of the drug overdose deaths and identifying high-risk areas are very useful to authorities, yet there are no interpretations provided to help them identify what the actual causes for the crisis are. The combination of crime and socioeconomic data makes the mission challenging because there are various factors for each category. In total, there are seven categories (crime rates, age demographics, education levels, economics (family and individual income), occupation types, poverty levels, and families receiving Supplemental and Nutrition Assistance Program (SNAP) benefits). Finding the category that contributes the most toward the crisis is crucial but challenging.

Algorithm 3 Drug Overdose mortalities explainability

Input: Crime-Socioeconomic CS_N^T
Output: Matrix of only one category Ranks $W_{M \times T}$, Matrix of without one category Ranks $O_{M \times T}$
Initialize: $t = \{1, ..., 7\}$, Intervals V_T, Models M_m

1: **for** the whole dataset of all categories CS_N^T **do**
2: **for** each index t get interval $V_t = i$ to k **do**
3: $Category =$ `Extract subset from` $CS_N^{i...k}$
4: $Rest =$`The remaining set` $CN_N^{T-Category}$
5: **for** each Model $m \in M$ **do**
6: $r1 \leftarrow$ Train m on $Category$
7: $r2 \leftarrow$ Train m on $Rest$
8: Add Results $r1$ to W_m^t
9: Add Results $r2$ to O_m^t
10: **end for**
11: **end for**
12: **end for**
13: **Sort** $W_{M \times T}$ and $O_{M \times T}$ based on rank
14: **Return** $W_{M \times T}$ and $O_{M \times T}$

The method proposed for explainability is displayed in algorithm 3. The main concept of the method is to train the models with only one category out of the seven categories and also train the models with all the categories except for the previously selected category. On one hand, training the models with only one category can underline the category's contribution toward the prediction and help intensify how each category performs among the rest of the categories. On the other hand, training the models by removing only one category each time can show the significance of the removed category when the models perform worse without the removed one. Finally, the two analyses will be sorted individually to find which category performed best across all models, the results are averaged to find which category has the most positive or negative impact on the models' prediction.

Specifically, the exhibited pseudo-code 3 takes as input the whole dataset CS_N^T where N indicates the number of counties and T is the number of categories. The output will be the two matrices $W_{M \times T}$ and $O_{M \times T}$ which include the sorted results from training the models where M annotates the number of models, and T is the number of categories. At the start, the algorithm initializes the intervals V_T. Each interval has the index of columns for each category in CS_N^T. For example to get the first category, the interval $V_{i...k}$ can be the first column i to column k. Lines $1-2$ start by including the whole dataset CS_N^T, then specifically going through each category of the dataset. In each interval, the data is divided into two subsets: $Category$ is only the subset of the chosen category, while $Rest$ is the subset of the remaining categories. In lines $5-9$, each model $m \in M$ trains the two subsets $Category$ and $Rest$, then stores the performance results in

$r1$ and $r2$ respectively. Further, $r1$ and $r2$ are added to the matrices $W_{M\times T}$ and $O_{M\times T}$ for later to be sorted. In line 13, each matrix is sorted by category across all the models. For instance, in $W_{M\times T}$, the $LinearRegression$ model has the results from each category trained individually. The best performing category obtains a rank of 1, the second performing category obtains 2, and so on. Once both matrices are returned from Alg 3, interpretations can be offered based on calculating the average of best-performing categories across all models in both matrices. Consequently, the analysis can conclude which of those categories (crime rates, age demographics, education levels, economics (family and individual income), occupation types, poverty levels, and families receiving (SNAP) benefits are the main driving factors for the drug overdose deaths crisis.

IV. Experiment

This section describes data collection, experimental settings, the subsequent results, and three case studies. The analysis of the results incorporates the effect of data augmentation on the accuracy of the prediction, the high-risk counties of drug overdose deaths ranked, and an explanation of the most contributing categories to the crisis.

A. Data Collection

The analysis in this research is based on three types of data collection for four years, 2017-2020: cumulative drug overdose deaths, crime reports data, and compiled socioeconomic statistics. The data has been collected for seven states, AZ, CA, FL, GA, NJ, NY, and VA, because of their data availability based on the county-level. Here are brief details about the datasets:

Drug overdosing death cases Dataset: For each state, the number of deaths has been collected, individually, from their official sources, which are the Department of Health of each state, except for California, where the data was collected from its official Overdose Surveillance Dashboard.

Crime Dataset: The collection varies from one state to another. The sources were from the AZ Department of Public Safety, Uniform Crime Reporting (UCR), Florida Department of Law Enforcement, NY Division of Criminal Justice, and states police departments.

Socioeconomic Dataset: As discussed in section II, other works studied specific subsets of socioeconomic factors that contribute the most to drug overdose, but we attempt to be comprehensive in our study by including all previous studies' categories and include more related categories based on recent drug reports. Table I briefly indicates the six socioeconomic categories collected from the U.S. Census Bureau.

B. Experiment Settings

In this research, there are three tasks (see sec III), and to evaluate those tasks, several experiments are implemented. For T1, four experiments are conducted for each type of data augmentation: (1) Dataset without augmentation (baseline) except for GCN, (2) County random shuffling, (3) County Feature distance, and (4) County Proximity. For T2, since the task is considered in two approaches (prediction and binary classification), we experiment with the widely used ML and DL models by utilizing the un-augmented dataset (as baselines) except for GCN, then comparing them with our results. Lastly, T3 is also evaluated by comparing our method's baselines suggested for T1 and T2. For experiments without augmentation, a random selection of 75-25 partitions is allocated between training-testing for all the models. A Grid-Search method is employed to fine-tune the model hyper-parameters. For the three augmented techniques experiments, the initial 500 original data points are used as the holdout for validation, and the rest of the augmented data are used for training. The GCN experiment has not been employed in a similar task before (according to our knowledge) and is only conducted on a non-augmented dataset as that is the only dataset based on which we can curate a spatial graph structure.

C. Experimental Results

A total of 6 experiments are illustrated in Tables II, III, and IV to validate the proposed framework (T1 and T2). Table II displays the results of predicting drug overdose death rates approach and comparing them to the baselines in terms of R-squared (R^2), Mean Absolute Error (MAE), and Mean Absolute Percent Error (MAPE). The baselines in this approach are the most widely applied ML and DL models using the unaugmented dataset. It is noticeable that the performance of the unaugmented dataset is not satisfactory across all the models in terms of R^2 and MAE except for GCN. This can be explained not only by the considerably small dataset of 500 but is exacerbated by the fact that it also has over 100 features which invariably affects the performance of the models. However, MLP gives an R^2 score of 0.63 and MAPE of 2.99, but if compared to the other best-performing augmentation methods in ML, the no augmentation method is still not superior except for GCN. GCN demonstrated competitive results over ML models even with augmentation methods. The transductive learning setting of the GCN allows there to be an aggregation of feature knowledge of the entire graph even though only a certain percentage of the nodes' labels are known during training. Furthermore, it also proves that the spatial aggregation method of GNN through neighborhood feature aggregation is particularly useful for this task which, to some extent, can overcome the problems faced by the paucity of data.

Random-Shuffling presents the least improvement on our base experimental settings for the augmentation methods. This can be attributed to the fact that the algorithm does not account for any spatial or feature vector-based correlation while creating new data points, which may result in instances that are the result of one or more noisy or outlying data, resulting in unreliable results. Nevertheless, as expected, both County Feature distance and County Proximity-based methods perform much better and are comparable to each other for certain models like MLP, LightGBM, or SVR. These two methods outperform other experimental methods because of the algorithm's choice of data points to be combined. Among the ML models, County Feature distance-based LightGBM

TABLE II: Experiment results

Model	No Augmentation			County Random Shuffling			County Feature Distance			County Proximity		
	R^2	MAE	$MAPE$	R^2	MAE	$MAPE$	R^2	MAE	$MAPE$	R^2	MAE	$MAPE$
Ridge	0.0481	0.6621	1.781	0.0981	0.6862	2.5593	0.3195	0.6028	1.9781	0.2541	0.6091	4.0246
Lasso	0.0510	0.6825	4.245	0.02619	0.7316	1.2010	0.1474	0.7709	1.511	0.1028	0.7493	1.9608
LightGBM	0.1314	0.7858	1.553	0.1129	0.7495	1.9972	**0.8039**	**0.4419**	**1.8412**	0.4672	0.7870	1.9624
SVR	0.0656	0.8202	1.8527	0.0918	0.7736	2.5740	0.4030	0.8039	2.1170	0.3719	0.8262	4.0246
XGBoost	0.0569	0.8402	1.984	0.0963	0.7642	1.9667	0.4813	0.8068	2.0252	0.6821	0.8689	2.3603
MLP	0.6356	1.0434	2.9892	0.0865	0.9039	4.0246	0.6388	0.9815	2.6573	**0.8389**	**0.8132**	**1.8293**
GCN	**0.8072**	**0.8131**	2.2364	—	—	—	—	—	—	—	—	—

TABLE III: Binary classification results based on unaugmented data (baselines)

Model	Accuracy	Percision	Recall	F1	AUC
Logistic	0.6120	0.6144	0.6035	0.5947	0.6580
KNN	**0.6217**	**0.6197**	0.5941	0.5910	0.6521
SVM	0.5984	0.5812	0.5984	0.5541	0.6570
MLP	0.5992	0.5885	0.6012	0.5847	0.6610
GCN	0.6082	0.5991	**0.6201**	**0.6118**	**0.6629**

TABLE IV: Binary classification results based on best performing augmentation method

Model	Accuracy	Percision	Recall	F1	AUC
Logistic	0.6916	0.6922	0.6916	0.6884	0.7642
KNN	**0.8737**	**0.8728**	**0.8734**	**0.8731**	**0.9319**
SVM	0.7047	0.7016	0.6947	0.7089	0.7874
MLP	0.7875	0.7873	0.7875	0.7873	0.8778

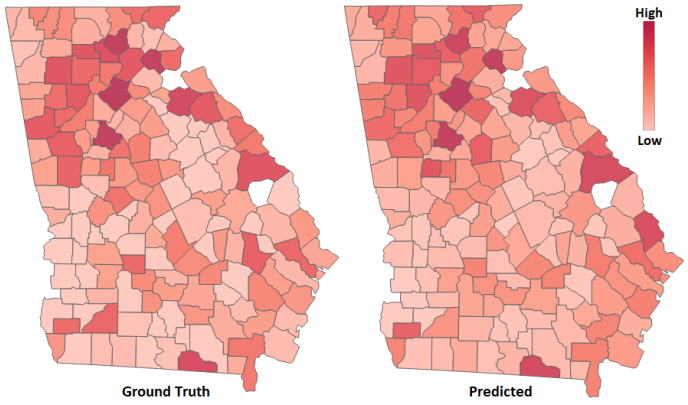

Fig. 2: 2017 Actual (left) and predicted (right) maps of GA counties indicating the severity of drug overdose deaths. The darker red color indicates more severity.

demonstrates its superiority over the rest. Meanwhile, County Proximity based MLP is the best performing DL approach for forecasting drug overdose death rates. Feature distance augmentation considers the farthest and closest County Feature distance to make new data, while the realistic spatial correlation between neighboring counties is also accounted for through the 'County Proximity' augmentation method. Therefore, we conclude that both methods are effective over the baselines, but each method has its own strengths and weaknesses.

Tables(III, IV) show the results of the classification approach. The baselines are considered in Table(III). We can see that the highest accuracy that any model achieves without augmentation is 0.6217, while the highest AUC was for GCN with a score of 0.6629. However, when we apply our best-performing augmentation method, we experience an increase ranging from 13% to 42% in all the metrics across the models. KNN outperforms both SVM and MLP, which were comparable for the non-augmented settings. We believe that it is strictly because of the County Feature distance algorithm's nature based on inter-county cosine feature distance. KNN also employs a similar distance measure to assign the nearest data points to the same label, which could be the reason for the significantly superior result for KNN compared to others. In summary, in both approaches (classification and prediction), our results outperformed the baselines to identify the highly impacted counties of drug overdose deaths.

D. Qualitative Analysis

The evaluation is also corroborated by three case studies for the states of GA, FL, and NY. Predicting the areas of high drug overdose death rates is accurately identified through our method when given a set of counties of any state, its crime reports data, and a set of socioeconomic features. The case studies are the following:

Case study 1 (Georgia): Figure 2 shows the ground truth map along with the predicted map of all the (154) counties in GA. The colors on the map indicate the gravity of drug overdose deaths. Our work not only predicted the most affected counties from the crisis in the same order of the ground truth map, such as Gwinnett, Lumpkin, Henry, Banks, Oglethorpe, Echols, and Dawson but also predicted the least affected counties, such as Chattooga and Colquitt.

Case study 2 (Florida):

In the case study of FL shown in figure 3 , the predicted map on the right and the left ground truth map barely show a few differences. The predicted drug overdose deaths most impacted counties in FL are Palm Beach, Duval, Brevard, Marion, Manatee, Lee, Sarasota, Gilchrist, and Nassau, in this order. The least impacted counties were Suwannee, DeSoto, and Charlotte. Gadsden county is one of the least impacted by the crisis, but our method did not identify it.

Case study 3 (New York): For the state of NY, there were more challenges than the rest of the states, as displayed in the comparison between the ground truth and predicted

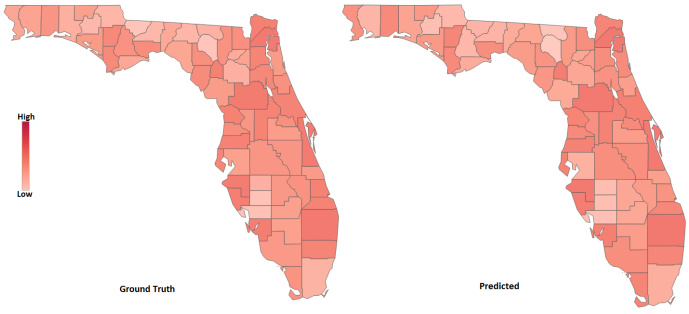

Fig. 3: 2017 Actual (left) and predicted (right) maps of FL counties indicating the severity of drug overdose deaths. The darker red color indicates more severity.

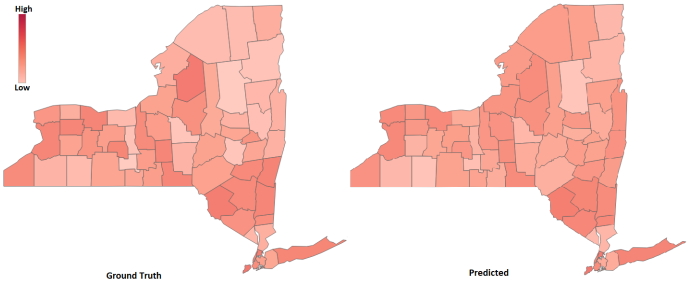

Fig. 4: 2017 Actual (left) and predicted (right) maps of NY counties indicating the severity of drug overdose deaths. The darker red color indicates more severity.

map (see figure 4). The most challenging counties that were not supposed to be identified as highly impacted were Columbia, Yates, and Genesee. On the other hand, Saratoga and Schoharie were misidentified among the least impacted counties. However, most of the high drug overdose death counties were predicted, such as Richmond, Lewis, Bronx, Sullivan, Monroe, Suffolk, and Cortland. Further, most of the least impacted counties were identified as well, such as Schuyler, Hamilton, Madison, Essex, and Seneca.

E. Explanations

Identifying the drug overdose deaths of the counties that are most and least affected with help from crime data and socioeconomic features is important, but at this point, there is not much clarity on what the most or least contributing categories are. Explainability can provide authorities feedback on which sector needs more reinforcement to mitigate the crisis.

The goal of **T3** is to rank which of the six categories accomplishes the best accuracy scores across all the models according to algorithm 3 in section III. The algorithm ranks the categories in each matrix $W_{M\times T}$ and $O_{M\times T}$ based on their accuracy scores by assigning a rank from 1 to 7 (1 is the highest rank). The top contributing features are in the Food Stamps/SNAP category, which achieved the best scores. The category includes features, e.g., "households not receiving food stamps at or above the poverty level," and features that measure the households that are in or out of the program. The second top contributing category is Education. This category's most correlated features are "high school graduate" or "bachelor's degree graduate." This category is followed by the crime rate category. The features of this category arranged according to their Pearson correlation are robbery, murder, aggravated assault, rape, motor vehicle thefts, larceny, and burglary. Then, the occupation category performs worse than the Crime category by a small margin but slightly better than the rest. The performance occupation category in $O_{M\times T}$ was the leverage over the last three categories.

Age demographics and Income categories were close in their ranks from one model to another. However, the Age demographics category contributes better in $W_{M\times T}$ but does not contribute as much as the Income category in $O_{M\times T}$. Finally, the Poverty category was expected to be the most contributing factor, but surprisingly, it was ranked last in both matrices. In summary, the application does not only provide predictions of high-risk counties of drug overdose deaths but also provides interpretable results for authorities to effectively mitigate the crisis in the short and long term.

V. Conclusion

This paper presents DOD-Explainer, an application that identifies the most and least drug overdosing impacted counties by utilizing an amalgamation of spatial crime reports and socioeconomic factors. It is challenging to work with drug mortality rate data to protect the privacy of families, but the study proposed solutions to data scarcity by implementing three augmentation methods. The application demonstrates its ability to forecast drug overdose mortality rates based on the data and classify the highly impacted areas based on the national threshold. The results show that the framework surpasses the baselines in both approaches, and the results are corroborated by three case studies. Finally, the paper proposes a method that provides explanations of the main contributing categories so authorities can have the clarity to take necessary actions on mitigating the crisis.

VI. Acknowledgement

This research is supported in part by National Science Foundation grants CNS-2141095. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, NSF, or the U.S. Government.

References

[1] D. Ciccarone, "The rise of illicit fentanyls, stimulants and the fourth wave of the opioid overdose crisis," *Current Opinion in Psychiatry*, vol. 34, no. 4, pp. 344–350, 2021.

[2] H. Kim and H. Yang, "Statistical analysis of county-level contributing factors to opioid-related overdose deaths in the united states," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 5860–5863, IEEE, 2020.

[3] A. Hollingsworth, C. J. Ruhm, and K. Simon, "Macroeconomic conditions and opioid abuse," *Journal of health economics*, vol. 56, pp. 222–233, 2017.

[4] V. A. Pear, W. R. Ponicki, A. Gaidus, K. M. Keyes, S. S. Martins, D. S. Fink, A. Rivera-Aguirre, P. J. Gruenewald, and M. Cerdá, "Urban-rural variation in the socioeconomic determinants of opioid overdose," *Drug and alcohol dependence*, vol. 195, pp. 66–73, 2019.

[5] M. J. Zoorob and J. L. Salemi, "Bowling alone, dying together: The role of social capital in mitigating the drug overdose epidemic in the united states," *Drug and alcohol dependence*, vol. 173, pp. 1–9, 2017.

[6] G. M. Heyman, N. McVicar, and H. Brownell, "Evidence that social-economic factors play an important role in drug overdose deaths," *International Journal of Drug Policy*, vol. 74, pp. 274–284, 2019.

[7] S. M. Monnat, "The contributions of socioeconomic and opioid supply factors to us drug mortality rates: Urban-rural and within-rural differences," *Journal of rural studies*, vol. 68, pp. 319–335, 2019.

[8] J. Programs, A. Craddock, J. J. Collins, and A. D. Timrots. U.S. Dept. of Justice, Office of Justice Programs, Bureau of Justice Statistics, 1994.

[9] C. Contreras and J. R. Hipp, "Drugs, crime, space, and time: A spatiotemporal examination of drug activity and crime rates," *Justice Quarterly*, vol. 37, no. 2, pp. 187–209, 2020.

[10] R. Rafaiee, S. Olyaee, and A. Sargolzaiee, "The relationship between the type of crime and drugs in addicted prisoners in zahedan central prison," *International journal of high risk behaviors & addiction*, vol. 2, no. 3, p. 139, 2013.

[11] A. K. Singh and D. Wu, "Sentiment analysis on substance use disorder (sud) tweets before and during covid-19 pandemic," in *International Conference on Human-Computer Interaction*, Springer, 2021.

[12] R. Sequeira, A. Gayen, N. Ganguly, S. K. Dandapat, and J. Chandra, "A large-scale study of the twitter follower network to characterize the spread of prescription drug abuse tweets," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1232–1244, 2019.

[13] J. A. Lossio-Ventura and J. Bian, "An inside look at the opioid crisis over twitter," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1496–1499, IEEE, 2018.

[14] K. M. Saifuddin, E. Akbas, M. Khanov, and J. Beaman, "Effects of covid-19 on individuals in opioid addiction recovery," in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1333–1340, IEEE, 2021.

[15] C. Buntain and J. Golbeck, "This is your twitter on drugs: Any questions?," in *Proceedings of the 24th international conference on World Wide Web*, pp. 777–782, 2015.

[16] E. Barenholtz, N. D. Fitzgerald, and W. E. Hahn, "Machine-learning approaches to substance-abuse research: Emerging trends and their implications," *Current opinion in psychiatry*, vol. 33, no. 4, 2020.

[17] H. Gong, C. Qian, Y. Wang, J. Yang, S. Yi, and Z. Xu, "Opioid abuse prediction based on multi-output support vector regression," in *Proceedings of the 2019 4th International Conference on Machine Learning Technologies*, pp. 36–41, 2019.

[18] K. M. Saifuddin, M. I. K. Islam, and E. Akbas, "Drug abuse detection in twitter-sphere: Graph-based approach," in *2021 IEEE International Conference on Big Data (Big Data)*, pp. 4136–4145, IEEE, 2021.

[19] K. G. McKim, C. Mai, D. Hess, and S. Niu, "Investigating drug addiction discourse on youtube," in *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, pp. 130–134, 2021.

[20] Y. Fan, Y. Zhang, Y. Ye, X. Li, and W. Zheng, "Social media for opioid addiction epidemiology: Automatic detection of opioid addicts from twitter and case studies," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017.

[21] D. Soni, T. Ghanem, B. Gomaa, and J. Schommer, "Leveraging twitter and neo4j to study the public use of opioids in the usa," in *Proceedings of the 2nd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics*, 2019.

[22] H. Hu, N. Phan, X. Ye, R. Jin, K. Ding, D. Dou, and H. T. Vo, "Drugtracker: A community-focused drug abuse monitoring and supporting system using social media and geospatial data (demo paper)," in *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2019.

[23] U. I. Islam, E. Haque, D. Alsalman, M. N. Islam, M. A. Moni, and I. H. Sarker, "A machine learning model for predicting individual substance abuse with associated risk-factors," *Annals of Data Science*, pp. 1–28, 2022.

[24] A. S. Wadekar, "A psychosocial approach to predicting substance use disorder (sud) among adolescents," in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 819–826, IEEE, 2020.

[25] J. P. Davis, P. Rao, B. Dilkina, J. Prindle, D. Eddie, N. C. Christie, G. DiGuiseppi, S. Saba, C. Ring, and M. Dennis, "Identifying individual and environmental predictors of opioid and psychostimulant use among adolescents and young adults following outpatient treatment," *Drug and alcohol dependence*, vol. 233, p. 109359, 2022.

[26] M. Tabar, H. Park, S. Winkler, D. Lee, A. Barman-Adhikari, and A. Yadav, "Identifying homeless youth at-risk of substance use disorder: Data-driven insights for policymakers," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3092–3100, 2020.

[27] X. Yang and J. Luo, "Tracking illicit drug dealing and abuse on instagram using multimodal analysis," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 4, pp. 1–15, 2017.

[28] A. M. Ertugrul, Y.-R. Lin, and T. Taskaya-Temizel, "Castnet: community-attentive spatio-temporal networks for opioid overdose forecasting," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 432–448, Springer, 2019.

[29] A. M. Ertugrul, Y.-R. Lin, C. Mair, and T. T. Temizel, "Forecasting heroin overdose occurrences from crime incidents," *Proceedings of SBP-BRiMS*, 2018.

[30] S. Mukherjee, N. Becker, W. Weeks, and J. L. Ferres, "Using internet search trends to forecast short term drug overdose deaths: A case study on connecticut," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1332–1339, IEEE, 2020.

[31] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Explainable ai in healthcare," in *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, pp. 1–2, IEEE, 2020.

[32] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "Gnn explainer: A tool for post-hoc explanation of graph neural networks," *arXiv preprint arXiv:1903.03894*, 2019.

[33] M. R. Karim, T. Döhmen, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, and S. Decker, "Deepcovidexplainer: explainable covid-19 diagnosis from chest x-ray images," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1034–1037, IEEE, 2020.

[34] A. Rodriguez, A. Tabassum, J. Cui, J. Xie, J. Ho, P. Agarwal, B. Adhikari, and B. A. Prakash, "Deepcovid: An operational deep learning-driven framework for explainable real-time covid-19 forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 15393–15400, 2021.

[35] Y. Li, H. Hu, and G. Zhou, "Using data augmentation in continuous authentication on smartphones," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 628–640, 2018.

[36] A. Le Guennec, S. Malinowski, and R. Tavenard, "Data augmentation for time series classification using convolutional neural networks," in *ECML/PKDD workshop on advanced analytics and learning on temporal data*, 2016.

[37] T. T. Um, F. M. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, and D. Kulić, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pp. 216–220, 2017.

[38] O. Steven Eyobu and D. S. Han, "Feature representation and data augmentation for human activity classification based on wearable imu sensor data using a deep lstm neural network," *Sensors*, vol. 18, no. 9, p. 2892, 2018.

[39] A. Alhamadani, S. Sarkar, L. Zhang, L. Alkulaib, and C.-T. Lu, "Forecasting high-risk areas of covid-19 infection through socioeconomic and static spatial analysis," in *2021 IEEE International Conference on Big Data (Big Data)*, pp. 4313–4322, 2021.

[40] R. E. Park and E. W. Burgess, *The city*. University of Chicago Press, 2019.

[41] R. L. Haffajee, L. A. Lin, A. S. Bohnert, and J. E. Goldstick, "Characteristics of us counties with high opioid overdose mortality and low capacity to deliver medications for opioid use disorder," *JAMA network open*, vol. 2, no. 6, pp. e196373–e196373, 2019.

[42] L. Zhao, J. Chen, F. Chen, F. Jin, W. Wang, C.-T. Lu, and N. Ramakrishnan, "Online flu epidemiological deep modeling on disease contact network," *GeoInformatica*, vol. 24, no. 2, pp. 443–475, 2020.

[43] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.